

GOODNESS-OF-FIT EVALUATION METHOD BETWEEN TWO SETS OF HOUSEHOLD MICRO-DATA FOR LAND-USE MICROSIMULATION MODELS

Noriko Otani, Faculty of Environmental and Information Studies, Tokyo City University, 3-3-1 Ushikubo-nishi, Tsuzuki, Yokohama, 224-8551 Japan, otani@tcu.ac.jp

Nao Sugiki, Research & Planning Department, Docon Co., Ltd., 4-1, 5-chome, 1-jo, Atsubetsuchuo, Atsubetsu, Sapporo, 004-8585 Japan, ns1491@docon.jp

Kazuaki Miyamoto, Faculty of Environmental and Information Studies, Tokyo City University, 3-3-1 Ushikubo-nishi, Tsuzuki, Yokohama, 224-8551 Japan, miyamoto@tcu.ac.jp

ABSTRACT

Land-use microsimulation models deal with detailed attributes of a household and its location which this study calls micro-data. However, less attention has been paid to the method to evaluate the goodness-of-fit between two sets of agent-based micro-data; an estimated and the observed sets of micro-data. At first, this study defines a distant measure between an estimated and the observed micro-data for each household. The goodness-of-fit is measured by the minimum sum of the distances for all households in the study area. Since the calculation cannot be conducted only with a conventional algorithm for an ordinary size of micro-data, an algorithm using GA, especially symbiotic evolution, is developed. The effectiveness of the method has been confirmed with the applications to the Person Trip Survey data in Sapporo Metropolitan Area.

Keywords: synthetic population, micro-data, goodness-of-fit, microsimulation, land-use model

1. INTRODUCTION

In the field of urban modeling, microsimulation is becoming a popular approach to describe the detailed changes in land use and transport in a metropolis. In this study, a set of data required by microsimulation models of households, in which a residential location model is a typical example, is selected for the subject.

*Goodness-of-Fit Evaluation Method between Two Sets of Household Micro-Data
for Land-Use Micro-Simulation Models*

OTANI, Noriko; SUGIKI, Nao; MIYAMOTO, Kazuaki

A household is characterized by many attributes such as number of members, ages of members, car ownership, housing type, location, etc. According to the principle of microsimulation, every necessary attribute of all households in a study area should be prepared for the base year of the simulation. However, it is not available, for the retrieval of individual micro-data from administrative registers is neither allowed nor desirable for privacy reasons. Therefore these models create “synthetic population” from general accessible aggregate data provided by the national census with additional information obtained by sample survey.

In most existing procedures of synthesizing population, the number of households by type is estimated after setting household types with IPF (iterative proportional fitting) method. However, this approach has several difficulties if the model has to deal with many kinds of attributes of households (Miyamoto, et al., 2009; Miyamoto, et al., 2010). Then another approach is to generate a set of individual households each of which has its unique set of attributes. This study call the set of attributes of an individual household, in general an individual agent, micro-data. Some attempts are being made to estimate a set of micro-data in a zone (Moeckel, et al., 2009; Miyamoto, et al., 2009). Although an observed set of micro-data is usually not available in a real application of microsimulation model, the validity of estimation method should be tested with a test data set. However, there is no method to evaluate the goodness-of-fit between such sets; estimated and observed agent-based micro-data sets.

The purpose of the present study is to establish an evaluation method of goodness-of-fit of an estimated set of micro-data with the observed one. The definition of the measure is rather naïve, and the measure can be obtained with a conventional calculation method so far as the number of agents is less than thirty or so. However, the number of calculations increases in proportion to the factorial ($N!$) of the number (N) of agents. The calculation cannot be conducted only with a conventional algorithm for an ordinary size of micro-data. Therefore, this study also develops an algorithm to make the measure feasible for a normal problem of micro-data. Although this evaluation method of goodness-of-fit is originally intended for the development of urban microsimulation models as mentioned above, it can be also applicable to the similar problem in other fields of study.

This paper first overviews state of the arts regarding goodness-of-fit evaluation measures used for synthesized population of microsimulation and the algorithm which is required to make the calculation feasible. Then, a measure is defined to evaluate goodness-of-fit between two sets of micro-data, followed by the algorithm which can be applied to find out the best approximate value of the measure with a certain number of agents. After confirming the performance of the method in a simple case with small number of agents, the method is applied to a micro-data set of 4000 households which are extracted from the Person Trip Survey data in Sapporo Metropolitan Area. Finally, the validity of the method is discussed with the result.

2. PREVIOUS WORK

2.1 Goodness-of-Fit of Cross-Classifying Tables

Pritchard and Miller (2009) discuss the goodness-of-fit of micro-data. Although they mention that the synthetic population could be tested for goodness-of-fit against the true population's characteristics, if complete "ground truth" were known, they don't indicate the way of goodness-of-fit test. With the absence of the ground truth, they refer to the test by cross-classifying the synthetic population to form one table which is obtained by IPF (Iterative Proportional Fitting)-based procedures.

If the household is classified by three categories (i, j, k), the synthetic population, can be compared to a validation table using the distance-based Standardized Root Mean Square Error (SRMSE) statistic (Knudsen and Fotheringham, 1986) ,

$$SRMSE = \frac{\sqrt{\frac{1}{IJK} \sum_{i,j,k} (\hat{N}_{ijk} - N_{ijk})^2}}{\frac{1}{IJK} \sum_{i,j,k} N_{ijk}} \quad (1.)$$

and a less value indicates a better fit. This statistic is calculated for each of the validation tables in turn and the goodness-of-fit statistics in each type are then averaged together to give an overall goodness-of-fit for the type. In this case, there is no problem in calculating the measure.

2.2 Calculating Goodness-of-Fit based on Symbiotic Evolution

Genetic algorithm (GA) is a popular technique to solve optimization problems imitating biological evolution process. Many types of GA methods have been proposed. Symbiotic evolution was proposed for forming Neural Networks (Moriarty, et al., 1995; Moriarty, et al., 1996; Moriarty, et al., 1998). It maintains two separate populations: a population of partial solutions and a population of whole solutions, and results in a fast, efficient genetic search and prevents convergence to suboptimal solutions. Its effectiveness had been demonstrated in not only Neural Networks but also Inductive Logic Programming and generating decision trees (Otani, et al., 2002; Otani, et al., 2004; Otani, et al., 2006).

We had defined the goodness-of-fit for micro-data with three attributes, each members' age, gender, and relationship with the head of household, and proposed the calculation algorithm based on symbiotic evolution (Otani, et al., 2009a; Otani, et al., 2009b). It was shown that the algorithm is effective by the experiments using practical data.

In this study, we propose the enhanced definition and algorithm for micro-data with more various types of attribute.

3. DEFINITION OF THE PROBLEM

3.1 Propositions

The problem is defined with the following propositions.

1. The subject is a set of agents, each of which has its multivariate attributes. In this study, it is a set of household micro-data in a study area.
2. Some attributes are continuous and the others are discrete. In this study, ages of household members are continuous. The gender of members, the relationship of each member with the head of household, housing type and the zone in which the household is located are discrete.
3. Member's genders and relationships with the head of household are categorized into the following 20 "general household member types".

head(male)	child1(male)	child2(male)	child3(male)
wife	child1(female)	child2(female)	child3(female)
head(female)	child's wife	brother	sister
grandchild(male)	grandchild(female)	father	mother
other1(male)	other2(male)	other1(female)	other2(female)

4. The types of housing are set up the following five categories.

owned detached house, rented detached house, owned condominium, rented apartment and others

5. The observed dataset with full information is available to test the validity of estimation methods, for example (Miyamoto, et al., 2009).
6. Estimated datasets are provided by several estimation methods.

The problem is to calculate the goodness-of-fit of each estimated dataset to decide which estimated dataset is more similar to the observed dataset.

3.2 Notations

In this study, attributes of micro-data are each member's age, each member's gender, each member's relationship with the head of household, housing type and residential zone.

Observed dataset B and estimated dataset E_j are denoted as follows:

$$B = \{ \{ \vec{a}_i = (a_{i1}, a_{i2}, \dots, a_{i20}), \vec{h}_i = (h_{i1}, h_{i2}), r_i \} | 1 \leq i \leq N \} \quad (2.)$$

$$E_j = \{ \{ \vec{a}_i^j = (a_{i1}^j, a_{i2}^j, \dots, a_{i20}^j), \vec{h}_i^j = (h_{i1}^j, h_{i2}^j), r_i^j \} | 1 \leq i \leq N \} \quad (3.)$$

*Goodness-of-Fit Evaluation Method between Two Sets of Household Micro-Data
for Land-Use Micro-Simulation Models*

OTANI, Noriko; SUGIKI, Nao; MIYAMOTO, Kazuaki

where N is the number of observed data. Each data in the both datasets is a set of two vectors \vec{a}_i^j (\vec{a}_i), \vec{h}_i^j (\vec{h}_i) and one scalar r_i^j (r_j). a_{ik} and a_{ik}^j are values in the genuine population and the j -th synthetic population respectively, denote the age of the member who ranks as the k -th household member type in the i -th household. In this paper, we call i "household ID". If nobody ranks as the k -th type, the number 999 is substituted as a dummy number for a_{ik} or a_{ik}^j to indicate the absence of the household member.

\vec{h}_i^j denotes the housing type of the i -th household. $h_{i1}^j = 0$ shows that they live in their own house, $h_{i1}^j = 1$ shows that they live in the rented accommodation. $h_{i2}^j = 0$ shows that they live in the detached house, $h_{i2}^j = 1$ shows that they live in the condominium or apartment. $\vec{h}_i^j = (2, 2)$ shows that they live in other type of housing.

r_i^j denotes the serial number of the residential zone of the i -th household.

For example, the data of i -th household with father(age:45), mother(age:42), son(age:15) and daughter(age:12), they live in their own detached house in zone No.3, is {(45, 15, 999, 999, 42, 12, 999,...), (0, 0), 3}.

4. GOONESS-OF-FIT INDICATOR

4.1 Definition

Generally, average and variance statistics are used to find the similarities between two sets; however, the present problem is not to match the two distributions. Since there is no indicator that evaluates the goodness-of-fit between micro-dataset, this study proposes using the minimum value of the normalized sum of three weighted distances, $edis(\vec{a}_i, \vec{a}_{\sigma(i)}^j)$, $mdis(\vec{h}_i, \vec{h}_{\sigma(i)}^j)$ and $tdis(r_i, r_{\sigma(i)}^j)$. $edis(\vec{a}_i, \vec{a}_{\sigma(i)}^j)$ is a distance in each member's age, each member's gender and member's relationship with the head of household. $mdis(\vec{h}_i, \vec{h}_{\sigma(i)}^j)$ is that in housing type. And, $tdis(r_i, r_{\sigma(i)}^j)$ is that in residential zone. The goodness-of-fit $Fit(E_j)$ of the estimated dataset E_j is defined as follows:

$$Fit(E_j) = \frac{\min_{\sigma \in S_N} \left\{ w_a \cdot \sum_{i=1}^N edis(\vec{a}_i, \vec{a}_{\sigma(i)}^j) + w_h \cdot \sum_{i=1}^N mdis(\vec{h}_i, \vec{h}_{\sigma(i)}^j) + w_r \cdot \sum_{i=1}^N tdis(r_i, r_{\sigma(i)}^j) \right\}}{N \cdot (w_a + w_h + w_r)} \quad (4.)$$

where S_N means the set of all bijections from the set $\{1, 2, \dots, N\}$ onto itself. A bijection σ is a function from a set $X = \{1, 2, \dots, N\}$ to a set $Y = \{1, 2, \dots, N\}$ with the property that, for every y in Y , there is exactly one x in X such that $\sigma(x) = y$ and no unmapped element exists in either X or Y . As S_N has $N!$ elements, it is necessary to calculate the sum $N!$ times for one estimated dataset. w_a , w_h , w_r are weights for age (gender, relationship with the head of household), housing type and residential zone respectively.

Normalized Euclidean distance is used for the distance $edis(\vec{a}_i, \vec{a}_{\sigma(i)}^j)$ between \vec{a}_i and $\vec{a}_{\sigma(i)}^j$. To avoid both memory overflow and the excessive influence of the vacant values, i.e., 999; the ceiling value of the squared difference between elements $DiffMax$ is used as follows:

*Goodness-of-Fit Evaluation Method between Two Sets of Household Micro-Data
for Land-Use Micro-Simulation Models*

OTANI, Noriko; SUGIKI, Nao; MIYAMOTO, Kazuaki

$$edis(\bar{a}_i, \bar{a}_{\sigma(i)}^j) = \sqrt{\frac{1}{20 \cdot DiffMax} \sum_{k=1}^{20} \min((a_{ik} - a_{\sigma(i)k}^j)^2, DiffMax)} \quad (5.)$$

Normalized Manhattan distance is used for the distance $mdis(\bar{h}_i, \bar{h}_{\sigma(i)}^j)$ between \bar{h}_i and $\bar{h}_{\sigma(i)}^j$.

$$mdis(\bar{h}_i, \bar{h}_{\sigma(i)}^j) = \frac{1}{4} \sum_{k=1}^2 |h_{ik} - h_{\sigma(i)k}^j| \quad (6.)$$

Normalized time-distance is used for the distance $tdis(r_i, r_{\sigma(i)}^j)$ between r_i and $r_{\sigma(i)}^j$.

$$tdis(r_i, r_{\sigma(i)}^j) = \frac{\text{time-distance between zone } r_i \text{ and zone } r_{\sigma(i)}^j}{\text{the maximum value of all time-distances}} \quad (7.)$$

4.2 Computational Complexity

Calculating the distance sum is not so complex. But as N grows, the number of elements in S_N becomes larger and calculating time increases rapidly. The combinatorial explosion problem occurs.

In calculating the goodness-of-fit, the minimum value is selected after calculating all distance sums. In other words, calculating the goodness-of-fit is same as searching the minimum value from a lot of distance sums. This is the combinatorial optimization where GA works effectively.

5. OPTIMIZATION ALGORITHM

5.1 Basic Idea

We propose the method using GA, especially symbiotic evolution for calculating the goodness-of-fit of E_j . In GA, candidate solutions to an optimization problem are called individuals, and their abstract representations are called chromosomes. GA improves a population of chromosomes through application of mutation, crossover and selection operators repetitively. This process is called evolution. In symbiotic evolution, two populations, a population of partial solutions and a population of whole solutions, are evolved. In our method, an individual in the partial solution population expresses the L_p pairs of household ID in A and E_j , and an individual in the whole solution population expresses L_w combinations of some partial solution individuals.

5.2 Computation Steps

The chromosome of an individual in the partial solution population is expressed as a bit sequence whose length is $L_p \times 32$ (see Figure 1). Every 16 bits means one binary number. $L_p \times 2$ binary numbers are transformed to decimal numbers that mean household (HH) ID.

*Goodness-of-Fit Evaluation Method between Two Sets of Household Micro-Data
for Land-Use Micro-Simulation Models
OTANI, Noriko; SUGIKI, Nao; MIYAMOTO, Kazuaki*

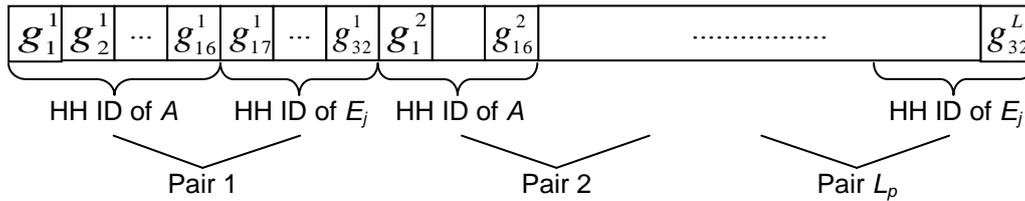


Figure 1 – Chromosome of an Individual in the Partial Solution Population

The chromosome of an individual in the whole solution population is expressed as $L_w = \lceil N / L_p \rceil$ pointers that point an individual in the partial solution population (see Figure 2). Algorithm for generating household ID pairs from an individual in the whole solution population is shown in

Figure 3. This algorithm can avoid making lethal genes that represent an inconsistent solution.

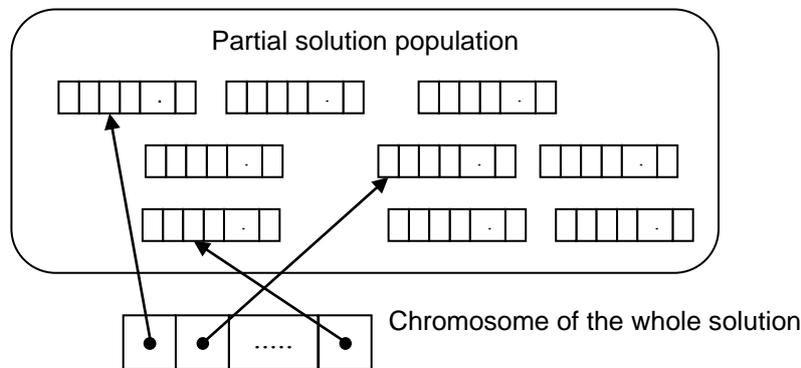


Figure 2 – Chromosome of an Individual in the Whole Solution Population

```

makeHHIDpairs() {
  for i:=1 to Lw {
    for j:=1 to Lp {
      fidA:=the household ID calculated gj1-gj16 pointed by the i-th pointer;
      fidP:=the household ID calculated gj17-gj32 pointed by the i-th pointer;
      if(both fidA and fidP are not used for making pairs) {
        make a pair (fidA, fidP);
      }
    }
  }
  for k:=1 to N {
    if(k is not used for making pairs as fidA) {
      fidP:=min(numbers that are not used for making pairs as fidP);
      make a pair (k, fidP);
    }
  }
}

```

Figure 3 – Algorithm for Generating HH ID Pairs from a Whole Solution Individual

The fitness value of an individual in the whole solution population is the distance summation calculated from household ID pairs. The smaller fitness value is the higher evaluation is.

*Goodness-of-Fit Evaluation Method between Two Sets of Household Micro-Data
for Land-Use Micro-Simulation Models*

OTANI, Noriko; SUGIKI, Nao; MIYAMOTO, Kazuaki

Individuals in the partial solution population are evaluated using the whole solution individuals each of which is pointing the partial solution individual. The partial solution individual is evaluated higher by being pointed from the better whole solution individual.

Both populations are evolved using the general GA operators; one-point crossover and mutation. One-point crossover operator makes two children by swapping all genes beyond one selected point between the two parent chromosomes. Mutation operator changes values in chromosomes at the probability of P_m . MGG generation alternation model, which is effective for avoiding the early convergence (Sato, et al., 1997), is applied.

The whole steps are as follows:

1. Make the initial partial solution population randomly.
2. Make the initial whole solution population randomly.
3. Evaluate whole solution individuals.
4. Evaluate partial solution individuals.
5. Evolve the partial solution population.
6. Evolve the whole solution population.
7. Repeat 1-6 for a pre-set number of times denoted by G .
8. Output the minimum fitness of the whole solution individual.

6. APPLICATIONS

6.1 Performance Test with Thorough Search

Performances between thorough search and proposed method were compared using small dataset with $N=14-19$. Parameters used in the proposed method are shown in Table 1 and three weights w_a , w_h , w_r are set to all 1. All experiments in this paper were conducted on a workstation with Intel Xeon 2.5GHz CPU and 32GB RAM.

In the all tests, the goodness-of-fit values outputted by the proposed method had been same as the thorough search. Times for computation of each test are shown in Table 2. While times of thorough search increase with N , times of proposed method have little change.

Table 1 – Parameters

Parameter name	Value
Number of individuals in a whole solution population	1000
Number of individuals in a partial solution population	1000
Mutation rate P_m	0.001
Repeat frequency G	5000
Length of partial solution chromosome L_p	5

Table 2 – Comparison of Time for Computation

N	Thorough search [sec]	Proposed [sec]	Goodness-of-fit
14	0.52	3.14	0.14142
15	2.25	3.35	0.13762
16	9.47	3.67	0.13663
17	10.59	3.91	0.12914
18	83.00	4.01	0.13109
19	394.83	4.12	0.12420

6.2 Performance Test with Practical Data

We prepared the observed dataset B ($N=4000$) of households in Sapporo metropolitan area. Here, values of age are transformed to the five-year age band to speed up the calculation. As it is impossible to know the goodness-of-fit for a large dataset, we prepared estimated datasets so that degrees of differences between them are evident. If we can get a large goodness-of-fit for an estimated dataset whose degree of difference from the observed data is high and a small goodness-of-fit for an estimated dataset whose degree of difference from the observed data is low, we can conclude that the proposed method is valid.

The estimated datasets E_{aj} are made as follows:

1. Select $j \times 400$ data from B randomly.
2. Subtract 5 from the age of one member selected randomly in the half of selected data.
3. Add 5 to the age of one member selected randomly in the rest of selected data.

The estimated datasets E_{bj} are made as follows:

1. Select $j \times 400$ data from B randomly.
2. In case of single-person household, change the householder's gender.
3. In case of multiple-person household, change the member type of one member selected randomly.

The estimated datasets E_{cj} are made as follows:

1. Select $j \times 400$ data from B randomly.
2. Reverse h_{i1}^j ($=0,1$) in the quarter of selected data.
3. Reverse h_{i2}^j ($=0,1$) in the next quarter of selected data.
4. Reverse h_{i1}^j, h_{i2}^j ($=0,1$) in the next quarter of selected data.
5. Change h_{i1}^j, h_{i2}^j ($=0,1$) into 2, and h_{i1}^j, h_{i2}^j ($=2$) into 0 or 1 in the rest of selected data.

The estimated datasets E_{dj} are made as follows:

1. Select $j \times 400$ data from B randomly.
2. Change the zone number randomly in the selected data.

*Goodness-of-Fit Evaluation Method between Two Sets of Household Micro-Data
for Land-Use Micro-Simulation Models*

OTANI, Noriko; SUGIKI, Nao; MIYAMOTO, Kazuaki

Note that the larger j , the larger difference from the observed dataset.

Parameters were same as Table 1 and three weights w_a , w_h , w_r are set to all 1. The goodness-of-fit of E_{a1} - E_{a5} , E_{b1} - E_{b5} , E_{c1} - E_{c5} , E_{d1} - E_{d5} were calculated by the proposed method 10 times. Average and standard deviation values of goodness-of-fit are shown in Table 3. In a very limited number of cases in which certain data happen to be altered quite largely, it is observed that the average of goodness-of-fit becomes worse than that of more altered data. However, in most cases, the goodness-of-fit, namely the minimum value of distance sum increases according to the number of altered data.

Table 3 – Average and Standard Deviation of Goodness-of-Fit

Data	Average	Standard deviation	Data	Average	Standard deviation
E_{a1}	0.20744	0.00045	E_{c1}	0.16741	0.00437
E_{a2}	0.20855	0.00027	E_{c2}	0.17822	0.00429
E_{a3}	0.20871	0.00038	E_{c3}	0.18916	0.00615
E_{a4}	0.21010	0.00020	E_{c4}	0.19849	0.00379
E_{a5}	0.21022	0.00032	E_{c5}	0.21132	0.00485
E_{b1}	0.21343	0.00017	E_{d1}	0.10421	0.00257
E_{b2}	0.21932	0.00047	E_{d2}	0.10737	0.00565
E_{b3}	0.22583	0.00045	E_{d3}	0.11243	0.01237
E_{b4}	0.23304	0.00034	E_{d4}	0.12121	0.00770
E_{b5}	0.23835	0.00018	E_{d5}	0.12123	0.01699

Next, it experimented to examine the change in goodness-of-fit by the difference of weights. The goodness-of-fit of E_{b1} - E_{b5} were calculated by the proposed method 10 times with two different weights sets respectively. Average values of goodness-of-fit are shown in Table 4. The estimated datasets E_{bj} are made by changing only member's type in the observed data. When w_a is smaller than other weights, the change in goodness-of-fit according to the number of altered data becomes small.

Table 4 – Average of Goodness-of-Fit for Different Weights

Data	$(w_a, w_h, w_r)=(1,2,1)$	$(w_a, w_h, w_r)=(1,1,2)$
E_{b1}	0.20990	0.22103
E_{b2}	0.21562	0.22587
E_{b3}	0.22104	0.23106
E_{b4}	0.22685	0.23626
E_{b5}	0.23170	0.24058

7. CONCLUDING REMARKS

The preposition of the study is that there are methods which synthesize individual households for microsimulation. A new method is proposed with a review of existing studies

*Goodness-of-Fit Evaluation Method between Two Sets of Household Micro-Data
for Land-Use Micro-Simulation Models*

OTANI, Noriko; SUGIKI, Nao; MIYAMOTO, Kazuaki

in Miyamoto, et al. (2010). However, existing studies have not tried to evaluate the goodness-of-fit on the individual level, for they assume observed dataset is not available. In this study, with the idea that goodness-of-fit on the individual level is necessary to evaluate the estimation or synthesize method, a measure of goodness-of-fit is originally proposed with the calculation method. Although more development is necessary to make the measure applicable to various attributes of micro-data which are used in exiting urban microsimulation models, a feasible measure to a simple case is proposed in this study. In addition, the measure may be also used in other fields of study than that of urban modeling.

ACKNOWLEDGEMENTS

This study was supported by Grant-in-Aid for Scientific Research (20360232) from Japan Society for the Promotion of Science.

REFERENCES

- Knudsen, D. C. and Fotheringham, A. S. (1986) Matrix Comparison, Goodness-of-Fit, and Spatial Interaction Modelling. *International Regional Science Review*, Vol. 10, No. 2, 127-147.
- Miyamoto, K., Kitazume, K. and Suzuki, A. (2008) A Review of Existing Urban Models in the World with emphasis on Theory, Function and Applications, Report of Grant-in-Aid for Scientific Research (18560524), Japan Society for the Promotion of Science (in Japanese).
- Miyamoto, K., and Sugiki, N., Kitazume, K., Suzuki, A. and Vichiensan, V., (2009) An Estimation Method of Household Micro-Data for the Base Year in Land-Use Micro Simulation, Proceedings of CUPUM'09.
- Miyamoto, K., Sugiki, N., Otani, N. and Vichiensan, V., (2010) An Agent Based Estimation Method of Household Micro-Data for the Base Year in Land-Use Microsimulation, Compendium of the 89th Transportation Research Board Annual Meeting
- Moeckel, R., Spiekermann, K., and Wegener, M. (2003) Creating a Synthetic Population, Proceedings of CUPUM '03, Sendai, CD-ROM.
- Moriarty, D. E. and Miikkulainen, R. (1995) Efficient Learning from Delayed Rewards through Symbiotic Evolution. Proceedings 12th International Conference on Machine Learning, 396-404.
- Moriarty, D. E. and Miikkulainen, R. (1996) Efficient Reinforcement Learning through Symbiotic Evolution. *Machine Learning*, Vol. 22, 11-32.
- Moriarty, D. E. and Miikkulainen, R. (1998) Hierarchical Evolution of Neural Networks. Proceedings IEEE World Congress on Computational Intelligence, 428-433.
- Otani, N. and Ohwada, H. (2002) Using Symbiotic Evolution to Improve the Estimated Accuracy for Inductive Logic Programming. *Journal of Japanese Society for Artificial Intelligence*, Vol. 17, No. 4, 431-438 (in Japanese).
- Otani, N. and Shimura, M. (2004) Generating the Simple Decision Tree with Symbiotic Evolution. *Journal of Japanese Society for Artificial Intelligence*, Vol. 19, No. 5, 399-404 (in Japanese).

*Goodness-of-Fit Evaluation Method between Two Sets of Household Micro-Data
for Land-Use Micro-Simulation Models*

OTANI, Noriko; SUGIKI, Nao; MIYAMOTO, Kazuaki

- Otani, N., Kaihara, M. and Shimura, M. (2006) Two Step Decision Tree for Polymer Discrimination. *Journal of Japanese Society for Artificial Intelligence*, Vol. 21, No. 3, 295-300 (in Japanese).
- Otani, N., Sugiki, N. and Miyamoto, K. (2009a) Goodness-of-Fit Evaluation Method between Observed and Estimated Sets of Micro-Data in Land-Use Micro-Simulation. *Proceedings of CUPUM'09*.
- Otani, N., Sugiki, N. and Miyamoto, K. (2009b) Goodness-of-Fit Evaluation Method between Micro-Data Sets Which Contains Discrete Attributes in Land-Use Micro-Simulation. *Proceedings of Infrastructure Planning, Japan Society of Civil Engineers* (in Japanese).
- Pritchard, D. R. and Miller, E.J. (2009) Advances in Agent Population Synthesis and Application in an Integrated Land Use / Transportation Model, 88th Annual Meeting Compendium of Papers, Transportation Research Board.
- Sato, H., Ono, I. and Kobayashi, S. (1997) A New Generation Alternation Model of Genetic Algorithms and Its Assessment. *Journal of Japanese Society for Artificial Intelligence*, Vol. 12, No. 5, 734-744 (in Japanese).
- Wegener, M. (2003) Overview of Land-Use Transport Models, *Proceedings of CUPUM'03*.